



# **CREDIT EDA ASSIGNMENT**

**BY**

**ANANTAKUMAAR V R**

**IIIT-B**

# Problem Statement - I

- ▶ In this assignment, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.
- ▶ The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history.
- ▶ Because of that, some consumers use it as their advantage by becoming a defaulter.
- ▶ Suppose you work for a consumer finance company which specializes in lending various types of loans to urban customers.
- ▶ You have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.
- ▶ In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

# Approach of EDA

- ▶ Importing Relevant Libraries
- ▶ Load the Data
- ▶ Data Structure Checking Methods:
  1. Describe
  2. Info
  3. Shape
  4. Counting nulls – Percentage Method
- ▶ Converting to datatypes
- ▶ Outlier values
- ▶ Visualizing and Merging the Data

# EDA METHODS

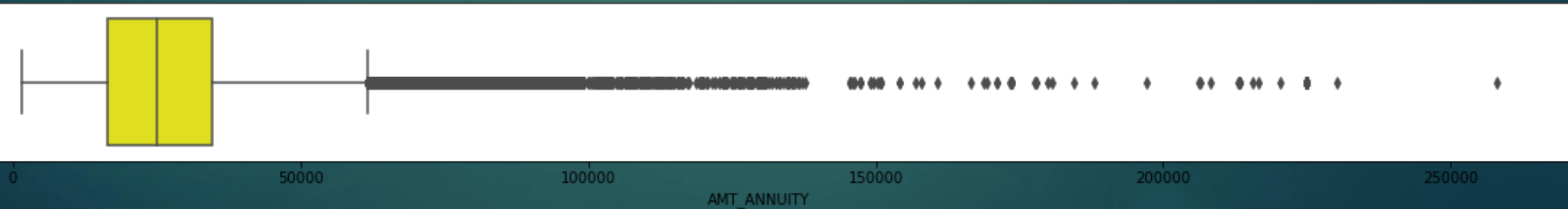
- ▶ Missing values in percentile calculation method
  1. Higher missing values than 40 % - removed
  2. Lower missing values – Replaced by zeroes
- ▶ Check datatypes of column for analysis
  1. Convert object to categorical type
- ▶ Checking the values of the outliers
  1. Box plot
- ▶ Analysis
- ▶ Dividing the target data into two sets consisting of 0 and 1
  1. Categorical analysis
  2. Correlation for numerical analysis.



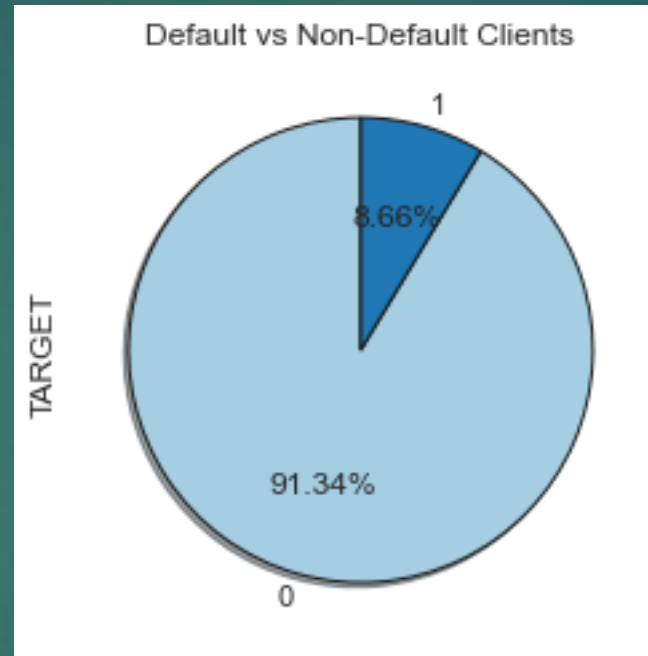
# INFERENCES AND VISUALIZATION OF DATA

## OUTLIERS OF AMT\_GOODS\_PRICE AND AMT\_ANNUITY

By inference, we can say that both the data contains outliers in a higher range.

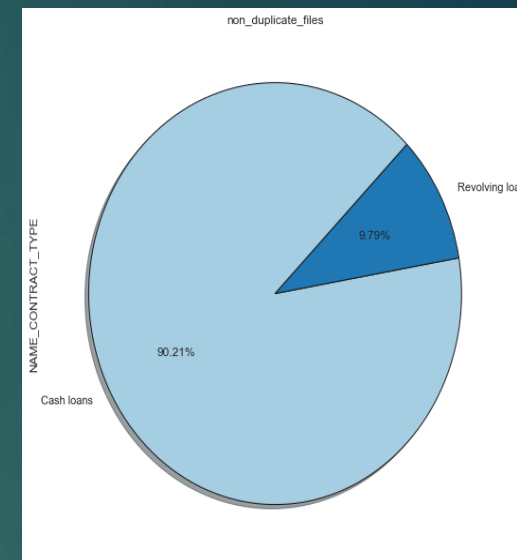
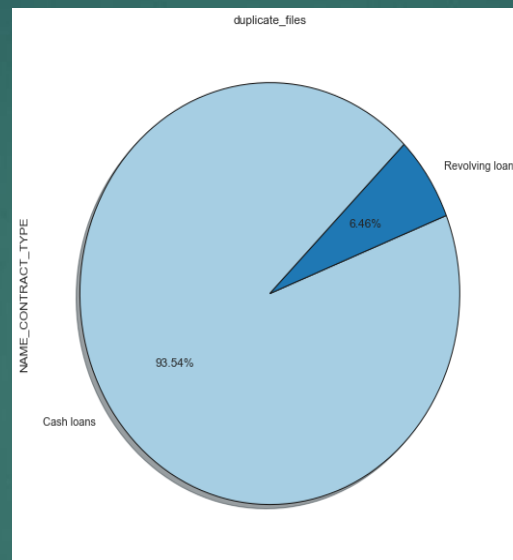
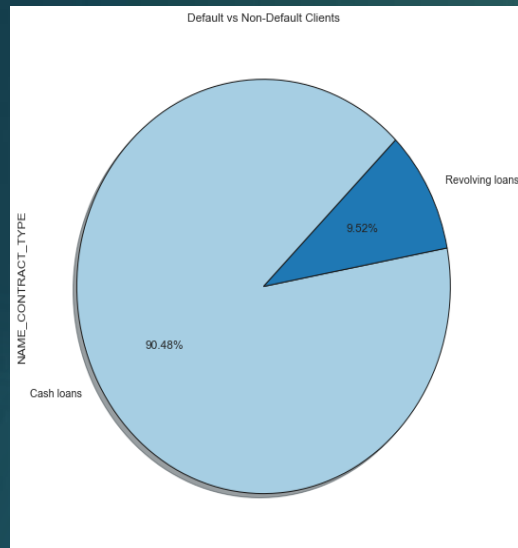


# CHEATERS VS USUAL BANK CUSTOMERS



- FROM THIS IMAGE, WE CAN INFER THAT 8.66 % OF THE BANK CUSTOMERS ARE FRAUDSTERS.

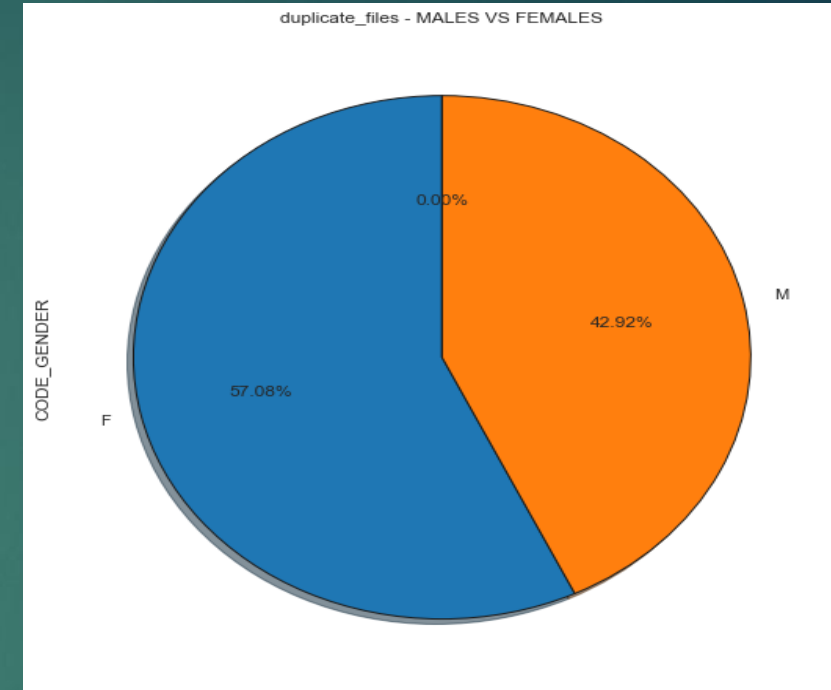
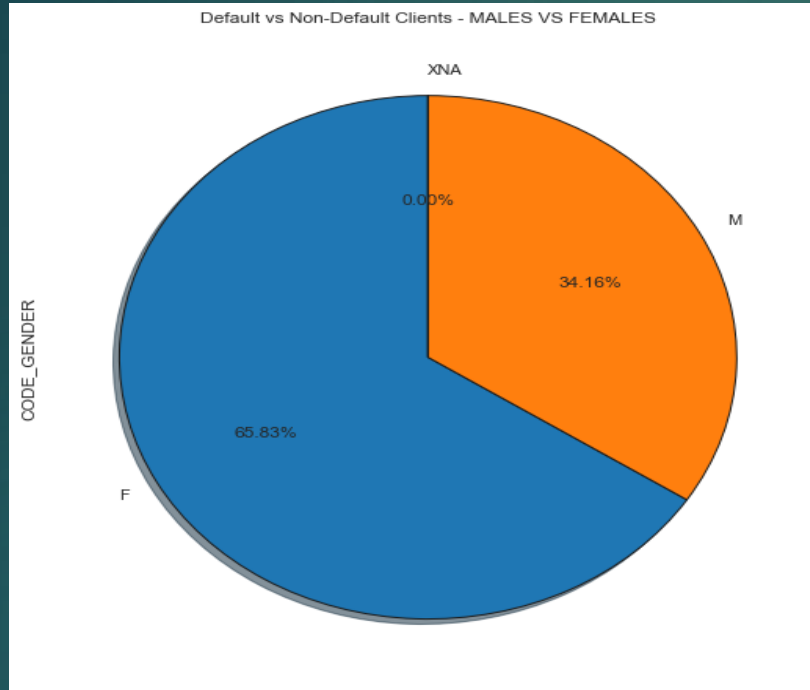
# IS CASH LOANS OR REVOLVING LOANS MORE POPULAR / DEFAULT ?



- WE CAN SEE THAT THE CASH LOANS ARE MORE POPULAR THAN REVOLVING LOANS.
- AND IT IS BETTER TO KEEP CASH LOANS AS DEFAULT

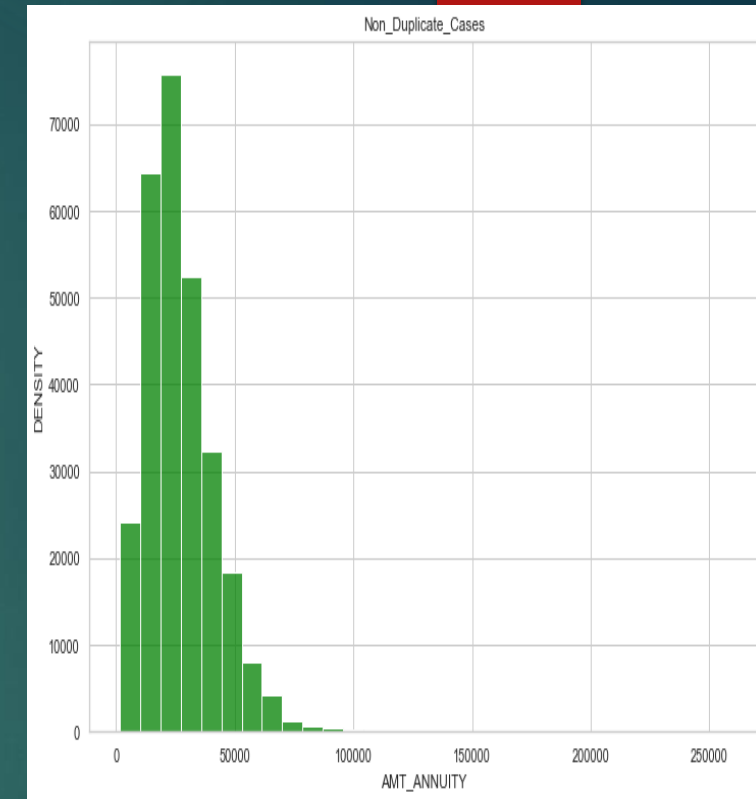
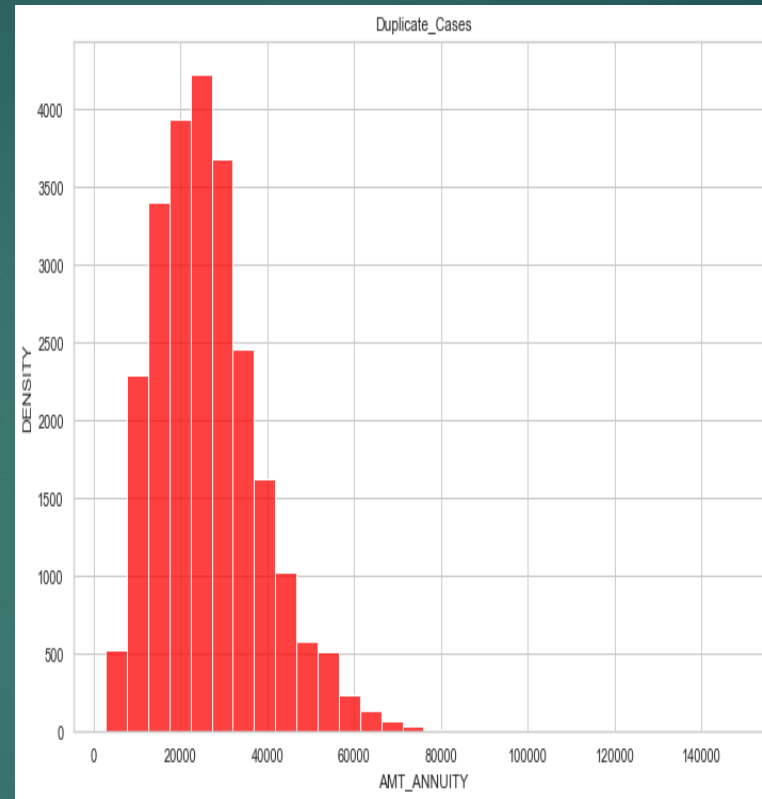
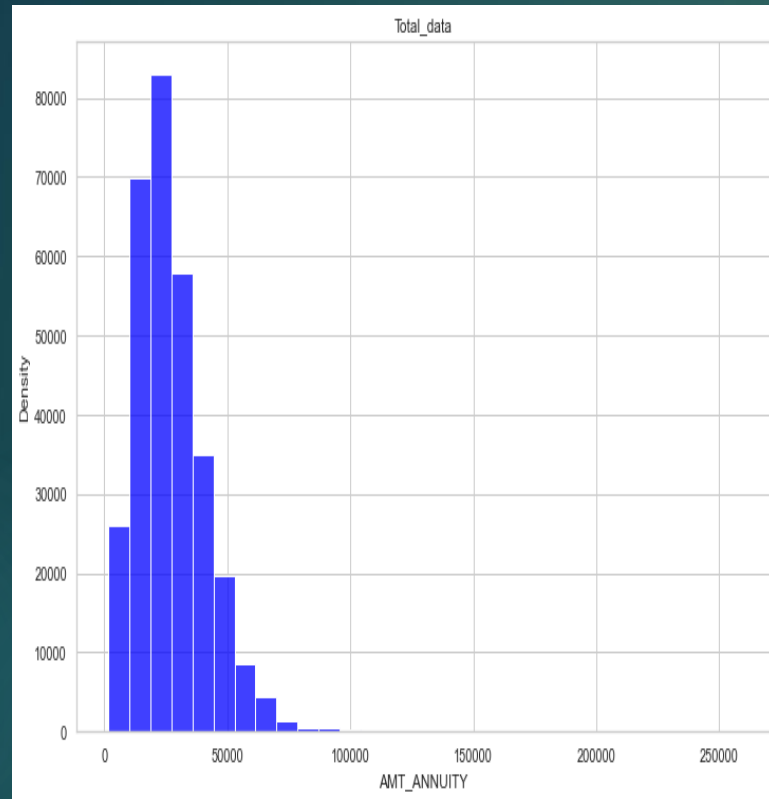


## WHICH GENDER HAS THE HIGHER PROBABILITY ?



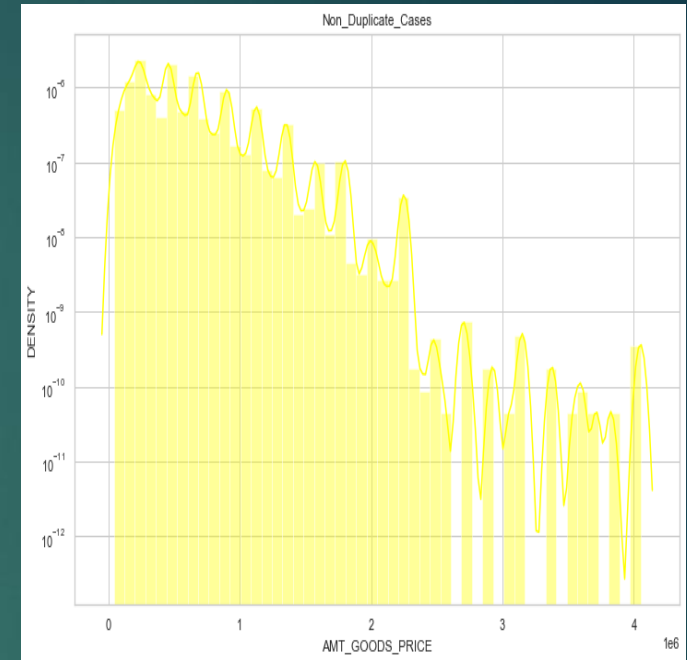
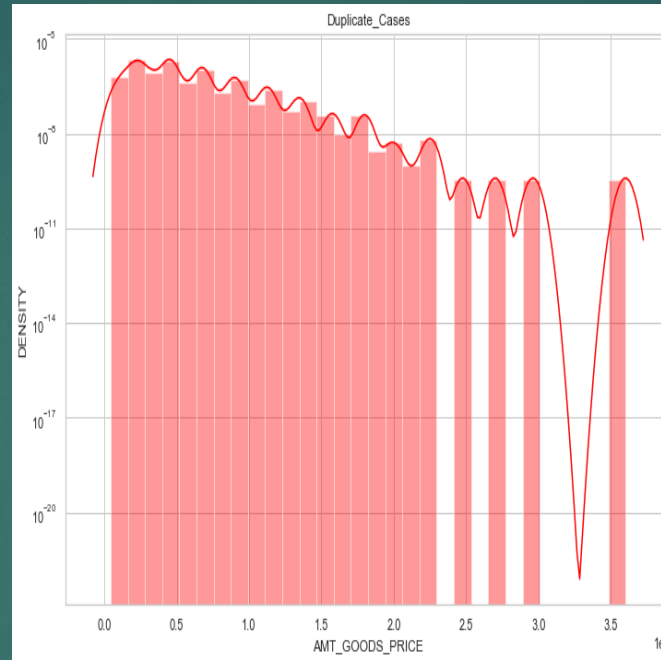
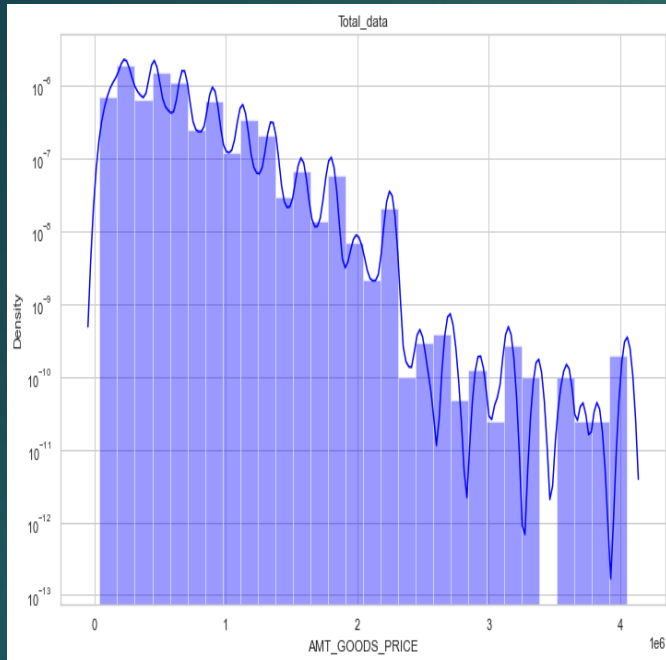
FROM THIS IMAGE WE CAN INFER THAT FEMALES ARE DEFAULTERS IN ALL CASES.

# RELATIONSHIP BETWEEN LOAN VALUE AND ANNUAL ANNUITY



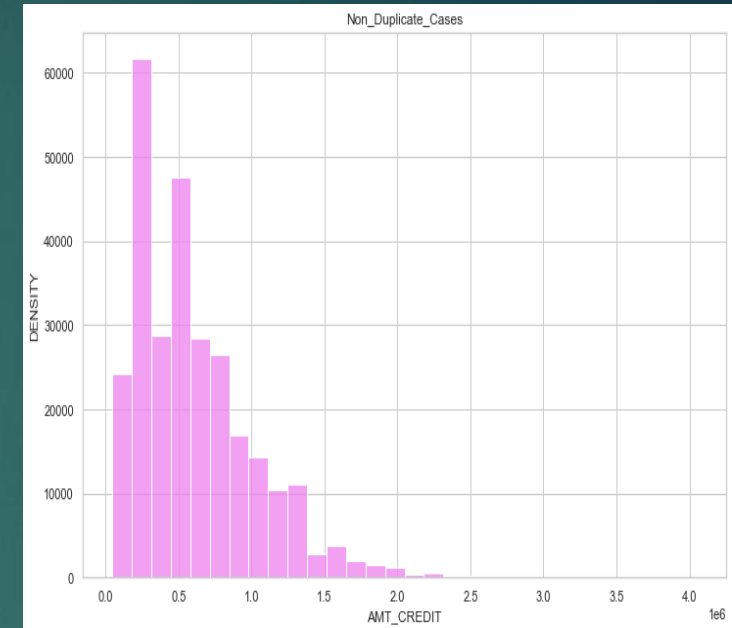
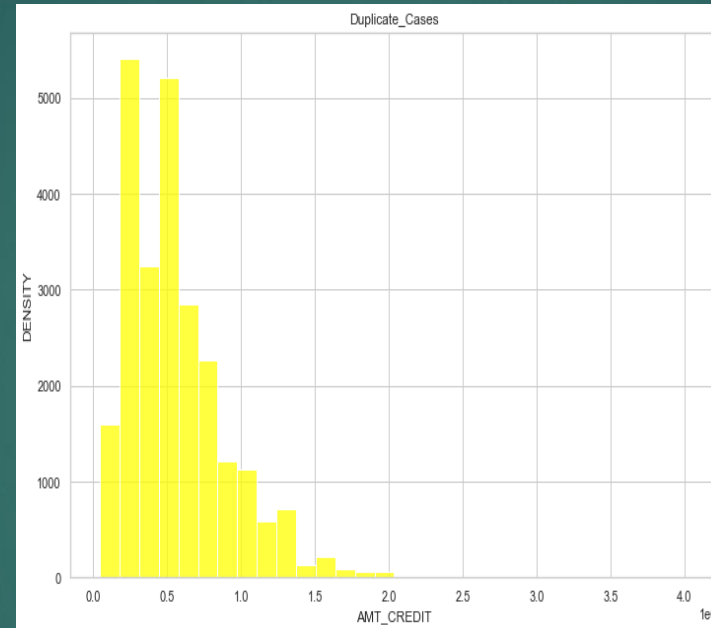
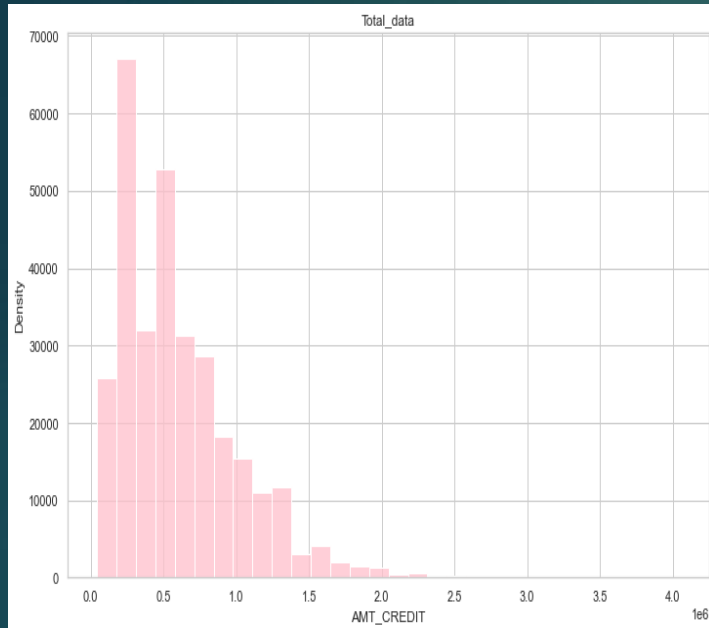
THE VOLUME OF LOAN VALUE IS HIGHER BUT VOLUME OF ANNUITY IS LOWER AS THE CUSTOMERS ARE NOT REPAYING THEIR LOAN AMOUNTS.

# GOODS PRICE



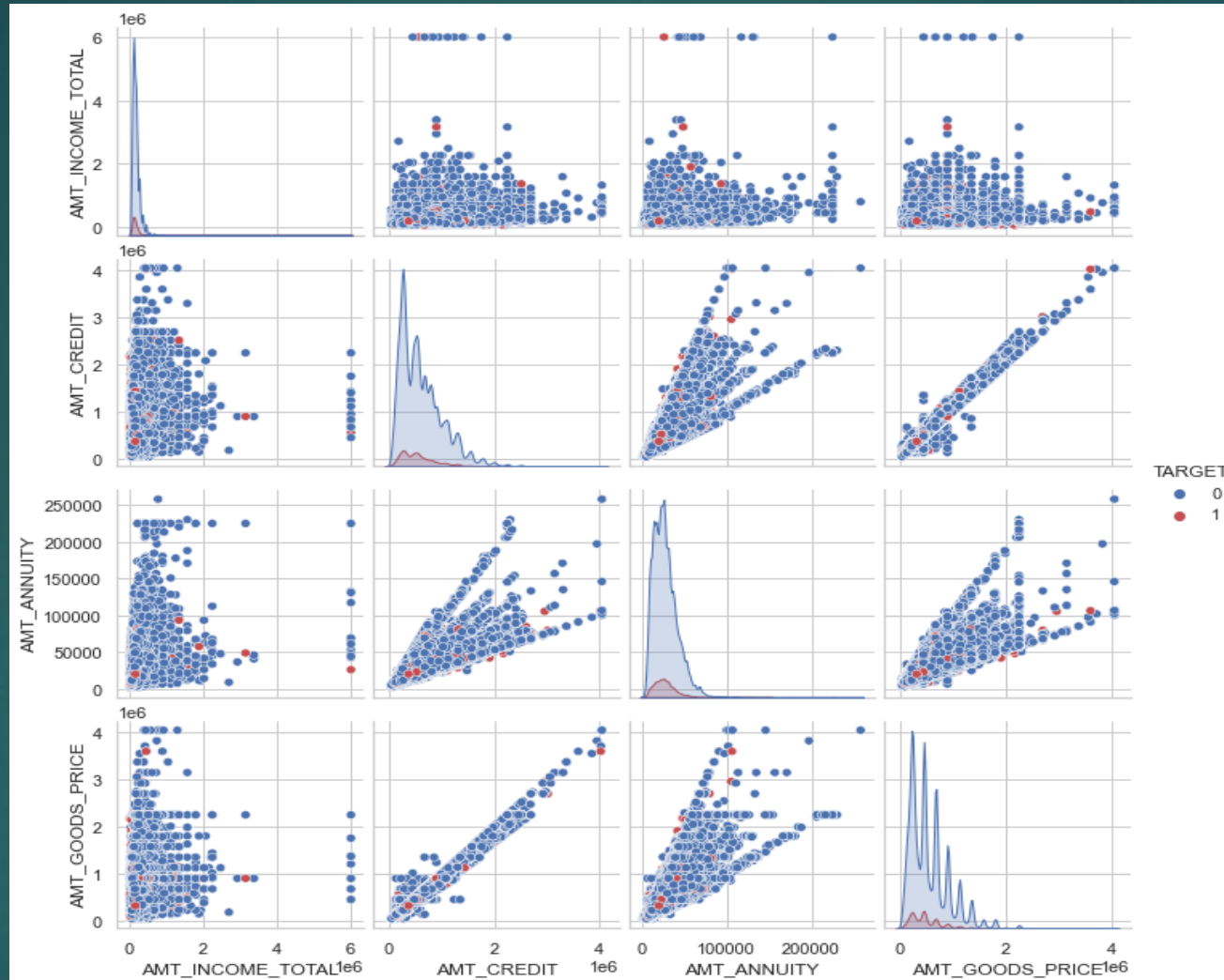
- AS FOR ALL THE DATA THE HIGHER RANGE OF MONEY SEEMS TO BE DECREASING.
- BECAUSE IF CUSTOMERS BUY A HIGHER LOAN THEN THEY'LL BY DEFAULT THEY PAY THEIR LOAN.

# THE RELATIONSHIP BETWEEN LOAN PAYERS AND LOAN PROVIDERS



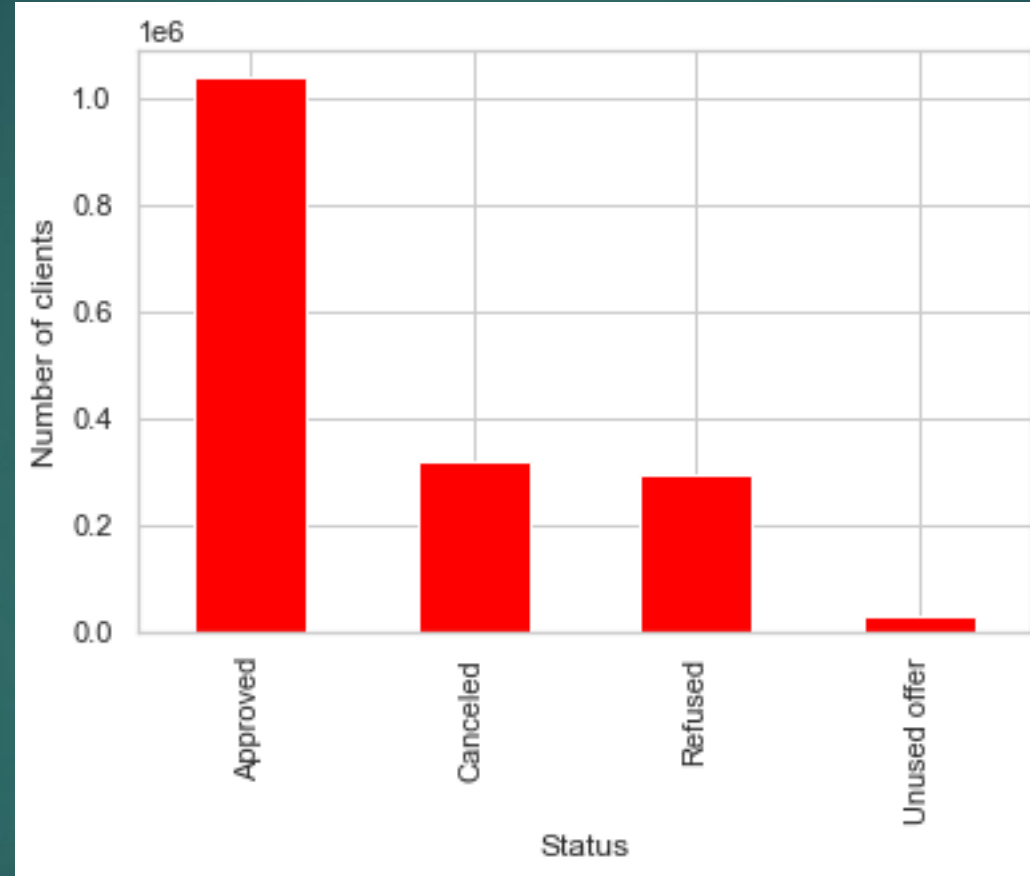
- AS THE BANK STARTED TO PROVIDE ITS LOAN AMOUNT ONLY AT HIGHER VALUE CUSTOMERS STARTED TO PAY THEIR LOANS.
- THAT IS THE REASON THE BANKS STOPPED TO PROVIDE LOW BANK LOANS FOR CUSTOMERS

# CORRELATION OF ALL AMOUNTS



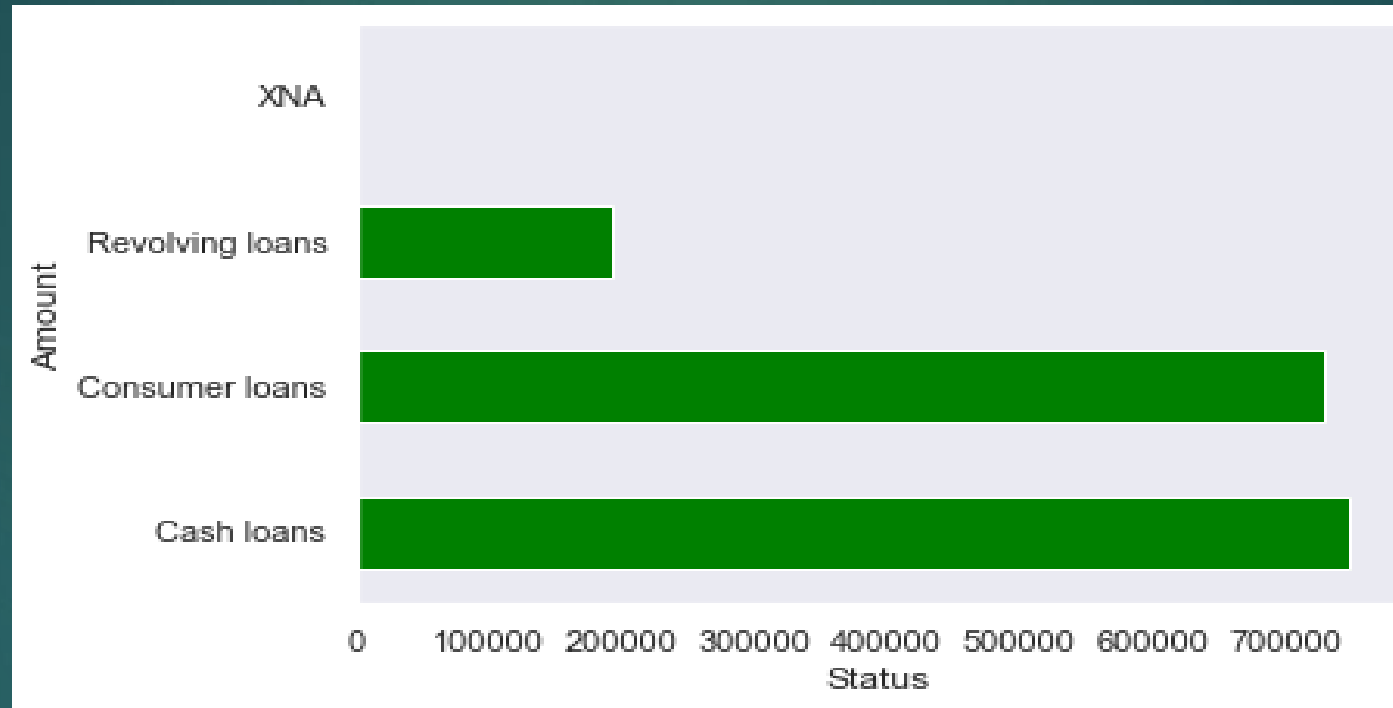
WE CAN CLEARLY INFER FROM DATA, THAT GOODS, ANNUITY, CREDIT AND INCOME HAS SOME HIGH CORRELATION VALUES. STILL THE RELATIONSHIP BETWEEN INCOME AND CREDIT HAS BEEN LOW DUE TO NON-PAYMENT OF LOANS FROM SOME CUSTOMERS.

## STATUS OF LOAN FOR CUSTOMERS



WE CAN CLEARLY INFER THAT FOR THE CUSTOMERS LOAN APPROVED IS AT HIGHER RANGE AND EVEN THOUGH SOME ARE BEING UNUSED BY SOME CUSTOMERS

# TYPES OF LOANS



IN THE TYPE OF LOANS CASH AND CONSUMER LOANS ARE ALMOST AT SIMILAR VALUES. WHEREAS REVOLVING LOANS HAS LOWER RECOGNITION.

# SUMMARY

- ▶ FROM THE DATA, ALMOST 91% CUSTOMERS ARE DEFAULT-LOAN PAYERS.
- ▶ CASH-LOANS AND CONSUMER-LOANS ARE SIMILAR IN DATA IN WHICH CASH-LOANS WILL BE REMAINED AS DEFAULTS.
- ▶ GENDER-WISE FEMALES WILL BE THE DEFAULTERS FOR PAYING THEIR LOAN AMOUNTS.
- ▶ BANKS WILL WILLINGLY PROVIDE HIGHER LOANS TO CUSTOMERS IN ORDER TO REDUCE THE RISK OF PROVIDING LOW LOANS TO CUSTOMERS TO WHICH THEY MAY NOT REPAY.
- ▶ IN CORRELATION POINT OF VIEW, GOODS PRICE SEEMS TO HAVE A HIGHER VALUE WHEN COMPARED TO ALL THE AMOUNTS.
- ▶ FROM THE DATA, WE CAN CONCLUDE THAT BANKS APPROVAL OF LOANS TO CUSTOMERS IS HIGHER AND IN A MUCH POSITIVE WAY.



**THANK YOU**